

Application of wavefield compressive sensing in surface wave tomography

Zhongwen Zhan,¹ Qingyang Li^{1,2} and Jianping Huang²

¹*Seismological Laboratory, Division of Geological and Planetary Sciences, California Institute of Technology, 1200 E. California Boulevard, Pasadena, CA 91125, USA. E-mail: zwzhan@gps.caltech.edu*

²*School of Geosciences, China University of Petroleum (East China), Qingdao 266580, China. E-mail: jphuang@upc.edu.cn*

Accepted 2018 February 27. Received 2018 February 8; in original form 2017 October 17

SUMMARY

Dense arrays allow sampling of seismic wavefield without significant aliasing, and surface wave tomography has benefitted from exploiting wavefield coherence among neighbouring stations. However, explicit or implicit assumptions about wavefield, irregular station spacing and noise still limit the applicability and resolution of current surface wave methods. Here, we propose to apply the theory of compressive sensing (CS) to seek a sparse representation of the surface wavefield using a plane-wave basis. Then we reconstruct the continuous surface wavefield on a dense regular grid before applying any tomographic methods. Synthetic tests demonstrate that wavefield CS improves robustness and resolution of Helmholtz tomography and wavefield gradiometry, especially when traditional approaches have difficulties due to sub-Nyquist sampling or complexities in wavefield.

Key words: Computational seismology; Crustal imaging; Seismic tomography; Surface waves and free oscillations; Wave propagation.

1 INTRODUCTION

In the last decade, we have witnessed the emergence of seismic arrays with a large number of sensors (i.e. large-N arrays), such as the USArray and the Nodal Seismic array at Long Beach, California, thanks to community-wide collaborative efforts and technological advances for portable instruments. As sensor technology continues to improve in the future, it will become possible or even routine to sample full seismic wavefield using larger and denser arrays. Exploiting these new data requires development of wavefield-based methods for Earth structure and earthquake source imaging. Several such methods (e.g. earthquake backprojection and detection, surface wave Helmholtz tomography) have already emerged and produced significant impacts on our understanding of the Earth (Ishii *et al.* 2005; Lin & Ritzwoller 2011; Yao *et al.* 2011; Ben-Zion *et al.* 2015; Hansen & Schmandt 2015; Inbal *et al.* 2016).

Seismic data from array sensors are discrete samples of spatially and temporally continuous wavefield. The temporal sampling is usually sufficiently even and dense (e.g. sampling frequency of 20–100 Hz) to avoid aliasing within the usual seismological period bands (frequency lower than 10 Hz), following the Nyquist–Shannon sampling theorem. However, the spatial sampling by arrays is usually not as dense or even as the temporal sampling, constrained by availability of sensors and where sensors can be deployed. Therefore, the Nyquist–Shannon sampling theorem cannot be applied here straightforwardly, and there will be spatial aliasing for seismic waves with half-wavelengths shorter than the average sensor spacing. Additionally, local noise may be strong and vary

from sensor to sensor, making it difficult to compare neighbouring records directly. How to reconstruct the original spatially continuous wavefield from our discrete and uneven samples with noise is a key question before further wavefield-based structure or source imaging.

Compressive sensing (CS) is a novel sampling paradigm in data acquisition (Candès & Wakin 2008; Davenport *et al.* 2011), and may apply to wavefield reconstruction from irregular spatial samples (Hindriks & Duijndam 1999; Zwartjes & Sacchi 2006; Herrmann *et al.* 2007; Herrmann & Hennenfent 2008; Hennenfent *et al.* 2010). CS theory asserts that certain signals can be recovered fully from relatively few measurements, provided that (1) the signal of interest can be sparsely represented under certain basis/frame Φ (i.e. sparsity), and (2) the sampling function Ψ is incoherent with the basis/frame Φ (i.e. incoherence, Candès & Wakin 2008). For example, among many sparsity-promoting methods, we can solve the following optimization problem

$$\min \{ \|\Psi\Phi m - d\|_2^2 + \lambda \|m\|_p \} \quad (1)$$

where the model vector m is a set of coefficients that sparsely represent the continuous wavefield to be sampled, s , by $s = \Phi m$, while $d = \Psi s + n$ is a vector of samples of s with noise n . The second term, $\lambda \|m\|_p$, imposes regularization over the coefficients m . Ideally, to promote sparsity of m , we would choose $p = 0$ (i.e. ℓ_0 minimization) that seeks the smallest number of non-zero values in m . However, this is impractical because ℓ_0 optimization problem is non-convex and generally impossible to solve. Instead, it is common

to choose $p = 1$ (i.e. ℓ_1 minimization) as a proxy for ℓ_0 minimization to seek a sparse solution (Donoho 2006; Candès *et al.* 2008; Davenport *et al.* 2011). This approximation has been shown to work remarkably well under many circumstances. Additionally, Candès *et al.* (2008) demonstrated that a sequence of iteratively reweighted ℓ_1 minimizations help to improve the quality of optimization. In this paper, we will only use the plain ℓ_1 minimization and leave the reweighted ℓ_1 minimization to a future discussion. Once the model vector m is estimated, then the wavefield s can be reconstructed as $s = \Phi m$, at any desired coordinates.

CS is becoming increasingly popular in many fields (e.g. compressive imaging, medical imaging, remote sensing and astronomy, Lustig *et al.* 2007; Bobin *et al.* 2008; Duarte *et al.* 2008; Herman & Strohmer 2009), because natural signals of interests often have proper bases/frames by which they can be sparsely represented. More specifically in geophysics, CS has been applied to seismic data acquisition/processing and source imaging. For example, Yao *et al.* (2011) used CS to image multiple point sources during large earthquakes with high-frequency teleseismic P waves on dense arrays. Herrmann *et al.* applied CS extensively in denoising and interpolating seismic profiles in exploration seismology problems (Herrmann *et al.* 2007; Herrmann & Hennenfent 2008; Hennenfent *et al.* 2010; Kumar *et al.* 2015), which share similar objectives as in this study but usually deal with small spatial scales and high-frequency body waves. Boehm *et al.* (2016) applied CS to overcome the bottleneck of massive memory requirements in full-waveform 3-D adjoint tomography. There are also wide interests in make seismic data acquisition more affordable using CS (Allegor *et al.* 2017; Baraniuk & Steeghs 2017; Kumar *et al.* 2017).

In this paper, we are essentially extending the CS approach to the scales of global and regional seismology, and focusing on its application to surface wave tomography. We aim to reconstruct surface wavefield on dense seismic arrays, to improve the robustness and resolution of structure imaging. To apply CS, first we need to find the proper basis/frame by which wavefield can be represented sparsely. There are many choices of basis/frame that are popular in exploration seismology, such as discrete cosine transform, curvelet (Candès & Demanet 2005), seislet (Fomel & Liu 2010) and wave atom (Demanet & Ying 2007). Some of these bases/frames (e.g. curvelet and wave atom) are constructed for purposes related to the seismic wave equation, thus features mathematically optimal sparsity under certain assumptions. For example, Candès & Demanet (2005) showed that the curvelet representation of wave propagators (or more generally, solutions to linear hyperbolic differential equations) is optimally sparse and well organized, as long as the medium is smooth or piecewisely smooth. This property makes curvelet an ideal choice of frame for bandlimited wavefield reconstruction in time domain (Herrmann & Hennenfent 2008). However, due to the dispersive nature of surface waves, we find that it is more natural to formulate the reconstruction problem in the frequency domain. In other words, we will reconstruct wavefields for individual frequencies, and estimate frequency-dependent phase velocity maps. We will leave the construction and inversion of dispersion curves for velocity structure as an independent problem, which can be solved by different methods (Ritzwoller *et al.* 2001, 2003; Yang *et al.* 2008; Haney & Tsai 2015), maybe jointly with receiver functions (Julia *et al.* 2000; Bodin *et al.* 2012; Shen *et al.* 2012). With this frequency-domain setup, we find that surface wavefield can be sparsely represented by a plane-wave basis (i.e. sparsity in the wavenumber domain). Meanwhile, the spatially random sampling function Ψ by seismic sensors is a network of Dirac functions, which is incoherent with the plane-wave basis. Therefore, wavefield

reconstruction with the plane-wave basis satisfies the two requirements of CS, sparsity and incoherence.

Practically, to handle the large data volume in most geophysical applications, we need an efficient ℓ_1 minimization solver to find the optimal reconstruction. Fortunately, many new and user-friendly tools have been developed in the applied mathematics community for problems in the form of eq. (1). Here, we will use the Templates for First-Order Conic Solvers (TFOCS) package because of its efficiency and robustness in our tests (Becker *et al.* 2011). TFOCS is library designed to construct efficient solvers for a variety of convex optimization problems, especially those in sparse signal recovery. Once the wavefield is reconstructed to a desired coordinates (e.g. a denser regular grid), we can apply any wavefield-based method with ease.

In the remainder of this paper, we will demonstrate how the CS approach works for surface wave problems on dense 2-D arrays. In Section 2, we will first briefly review current methods of wavefield-based surface wave tomography, their explicit or implicit assumptions and strengths/weaknesses. Then in Section 3, we will present surface wavefield reconstruction using CS and apply the current tomography methods to the reconstructed wavefield. In Section 4, we will demonstrate the effectiveness of CS in surface wave tomography by several synthetic examples, especially when traditional approaches have difficulties due to sub-Nyquist sampling or complexities in wavefield (e.g. strong multipathing and multiple interfering sources).

2 WAVEFIELD-BASED SURFACE WAVE TOMOGRAPHY

In the last decade, surface wave tomography methods based on either earthquake data or ambient noise interferometry have made enormous progresses in imaging shallow Earth structures. In particular, deployment of the USArray triggered a series of new wavefield-based methods, bringing unprecedented resolution to crustal and upper-mantle structures across the contiguous US (Langston 2007b; Liang & Langston 2009; Lin *et al.* 2009, 2013; Lin & Ritzwoller 2011; Ekström 2014, 2017; de Ridder & Biondi 2015; Jin & Gaherty 2015; Liu & Holt 2015; Bao *et al.* 2016). These wavefield-based methods work well partly because the propagation of a narrow-band single-mode surface wave approximately follows the 2-D wave equation (Tanimoto 1990; Tromp & Dahlen 1993), easily sampled by dense arrays:

$$\frac{1}{c(\vec{r}, \omega)^2} \frac{\partial^2 u(\vec{r}, t)}{\partial t^2} = \nabla^2 u(\vec{r}, t) \quad (2)$$

or equivalently in the frequency domain,

$$\frac{(i\omega)^2}{c(\vec{r}, \omega)^2} U(\vec{r}, \omega) = \nabla^2 U(\vec{r}, \omega) \quad (3)$$

where $c(\vec{r}, \omega)$ is the phase velocity distribution for the centre angular frequency ω , and $u(\vec{r}, t)$ or $U(\vec{r}, \omega)$ is the 2-D surface wave potential (in the time or the frequency domains, respectively), which for Rayleigh waves can be related to vertical displacement by a local amplification term (Tromp & Dahlen 1992; Lin *et al.* 2012). Here, we assume that the local amplification terms are negligible, similar to the assumptions made by Lin & Ritzwoller (2011) and de Ridder & Biondi (2015). If we further assume that $c(\vec{r}, \omega)$ is sufficiently smooth, then we can obtain the Helmholtz equation

$$\frac{1}{c(\vec{r}, \omega)^2} = |\nabla \tau(\vec{r})|^2 - \frac{\nabla^2 A(\vec{r})}{A(\vec{r}) \omega^2} \quad (4)$$

where $\tau(\vec{r})$ and $A(\vec{r})$ are the traveltime and amplitude fields, respectively. When the last term about amplitudes can be neglected (e.g. high-frequency approximation), we obtain the eikonal equation (Wielandt 1993; Lin *et al.* 2009). Because both traveltimes and amplitudes can be measured on 2-D dense arrays for individual earthquakes or virtual sources, we can apply eq. (4) to conduct eikonal or Helmholtz tomography, and then average the phase velocity maps from multiple sources for the final phase velocity map (Lin *et al.* 2009; Lin & Ritzwoller 2011). Making local plane-wave approximation to the solution of the wave equation (2) around a few nearby stations, Langston (2007b) developed the wave gradiometry to extract phase velocity, wave directionality, geometrical spreading and radiation pattern from the spatial and temporal derivatives of displacement, in the time or the frequency domains (Langston 2007a, b). Liu & Holt (2015) showed the connection between wave gradiometry and Helmholtz tomography. When station density is enough for us to sample the wavefield without spatial aliasing (i.e. station spacing $<$ half-wavelength), we can avoid most of the approximations above and return to the wave equations (2) and (3) to solve for the phase velocity directly (de Ridder & Biondi 2015), which we call wavefield gradiometry (WG) following de Ridder & Curtis (2016) hereafter. de Ridder & Biondi (2015) even showed that WG can be applied to raw ambient noise field consisting of diffuse surface waves from multiple sources/scatterers.

Helmholtz/Eikonal tomography, wave gradiometry and WG all take advantages of wavefield information sampled by dense arrays. Therefore, we classify them as wavefield-based surface wave tomography. More specifically, while time derivatives are relatively simple to compute for seismograms, wavefield-based methods all involve calculation of spatial derivatives of displacement or other attributes (e.g. traveltime and amplitude), enabled by dense arrays. For example, in Helmholtz tomography, spatial derivatives of traveltimes and amplitudes need to be computed for individual earthquakes or virtual sources (Lin & Ritzwoller 2011). In WG, the second-order spatial derivatives of displacement are computed using finite-difference (FD) approximations (de Ridder & Biondi 2015). These lead to two major practical considerations of wavefield-based methods: (1) FD approximation of derivatives on irregularly spaced stations, (2) sensitivity to incoherence or noise at neighbouring stations. To mitigate these difficulties, many pre-processing or post-processing methods have been used, such as smoothing of the fields, stacking/averaging over multiple earthquakes or virtual sources and calibration of the irregular FD stencil (Langston 2007b; Lin & Ritzwoller 2011; de Ridder & Curtis 2016). These ‘ad hoc’ methods often introduce new implicit assumptions and significantly affect the quality of tomography models.

From the point view of wavefield sampling, these difficulties are fundamentally due to operations applied to samples of the wavefield, not the true wavefield. The ‘ad hoc’ treatments are actually explicit or implicit assumptions about wavefield reconstructions. For example, the minimum-curvature smoothing that is often applied in Helmholtz tomography is equivalent to assuming that traveltime and amplitude fields of surface waves have certain levels of smoothness (Lin *et al.* 2009; Lin & Ritzwoller 2011). The local plane-wave approximation made in wave gradiometry assumes that the phase velocity is homogeneous locally within the subarray (Langston 2007a,b; Liang & Langston 2009). In WG, there are relatively few assumptions, but the sampling points need to be regular and dense enough so that the FD stencil is accurate, at least not biased systematically (de Ridder & Biondi 2015; de Ridder & Curtis 2016). All these assumptions about wavefield stabilize the solutions, but also limit the applications or accuracies. For exam-

ple, complex wavefield with strong multipathing are often difficult to analyse, and long duration of data or large number of events are often required to suppress noise and artefacts.

3 SURFACE WAVEFIELD RECONSTRUCTION

In this section, we will present a new wavefield reconstruction method for surface wave tomography, using the theory of CS as reviewed briefly in Introduction. Because the surface wave tomography referred here is to invert for phase velocity maps for single frequencies, we decide to formulate the wavefield reconstruction problem in the frequency domain. As in any of the wavefield-based methods, there are assumptions involved in reconstruction. Here, we assume that surface wavefield for individual frequencies can be sparsely represented by a plane-wave basis, or equivalently, sparse in the wavenumber \vec{k} domain. Although there is no mathematical proof supporting this assumption, we will demonstrate its effectiveness in Section 4 with synthetic examples. However, this intuition about sparsity is not new in surface wave seismology. Teleseismic surface waves distorted by multipathing and scattering are sometimes approximated by the sum of two plane waves with independent amplitudes, phases and propagation directions (Forsyth *et al.* 1998; Yang & Forsyth 2006). This would correspond to a sparse solution with only two non-zero plane-wave coefficients in our approach.

Taking Rayleigh waves as an example. We first Fourier transform vertical-component seismic waveforms $u(\vec{x}_i, t)$ to the frequency domain as complex-valued $U(\vec{x}_i, \omega)$, where \vec{x}_i are station locations. Then for each frequency ω , we conduct an optimization

$$\operatorname{argmin} \left\{ \|w\Psi\Phi m - wU(\vec{x}_i, \omega)\|_2^2 + \lambda \|m\|_1 \right\} \quad (5)$$

where m is a vector of plane-wave coefficients, Ψ and Φ are sampling function and plane-wave basis written in matrix form, respectively (Candès *et al.* 2006). w is a weighting term applied to both data and prediction. Because the plane-wave basis is non-local and periodic in space, we set w to be a spatial Hann taper function to avoid edge effects. This leads to worse wavefield reconstruction near edges of the data domain with lower weighting. In the future, using spatially localized basis/frame (e.g. wavelet and curvelet) may help improve reconstruction near edges. We solve this convex ℓ_1 optimization problem (eq. 5) using the TFOCS package to achieve a sparse \vec{k} -domain representation of $U(\vec{x}_i, \omega)$. The optimal damping parameter λ can be chosen by the L-curve approach for the best compromise between model complexity (more accurately, sparsity) and data misfit (Hansen & O’Leary 1993). Finally, we reconstruct the complete wavefield (e.g. on a denser, regular grid) $\tilde{U}(\vec{x}, \omega) = \Phi \tilde{m}$, where \tilde{m} is the estimated plane-wave coefficients.

As a simple synthetic example to demonstrate the procedures, Fig. 1(a) shows a smooth non-dispersive checkerboard phase velocity model, and we simulated the surface wavefield from an earthquake located to the southwest of the data domain by a 2-D FD method (Fig. 1a). Grid size of 12 km and time step of 1 s are used to ensure numerical stability for periods down to 10 s. The data domain has a dimension of 1536 km \times 1536 km, similar to the USArray footprint, and the input velocity model has an average velocity of 4 km s $^{-1}$ and ± 2 per cent maximum perturbations. Fig. 1(b) shows a snapshot of the wavefield at time = 450s, with the amplitude near the wave front slightly modulated by the velocity anomalies. We then transform the seismogram $U(\vec{x}, t)$ at any point \vec{x} to the frequency domain as $U(\vec{x}, \omega)$, which in the following we simply write as $U(\vec{x})$ for any fixed frequency. The real part of synthetic

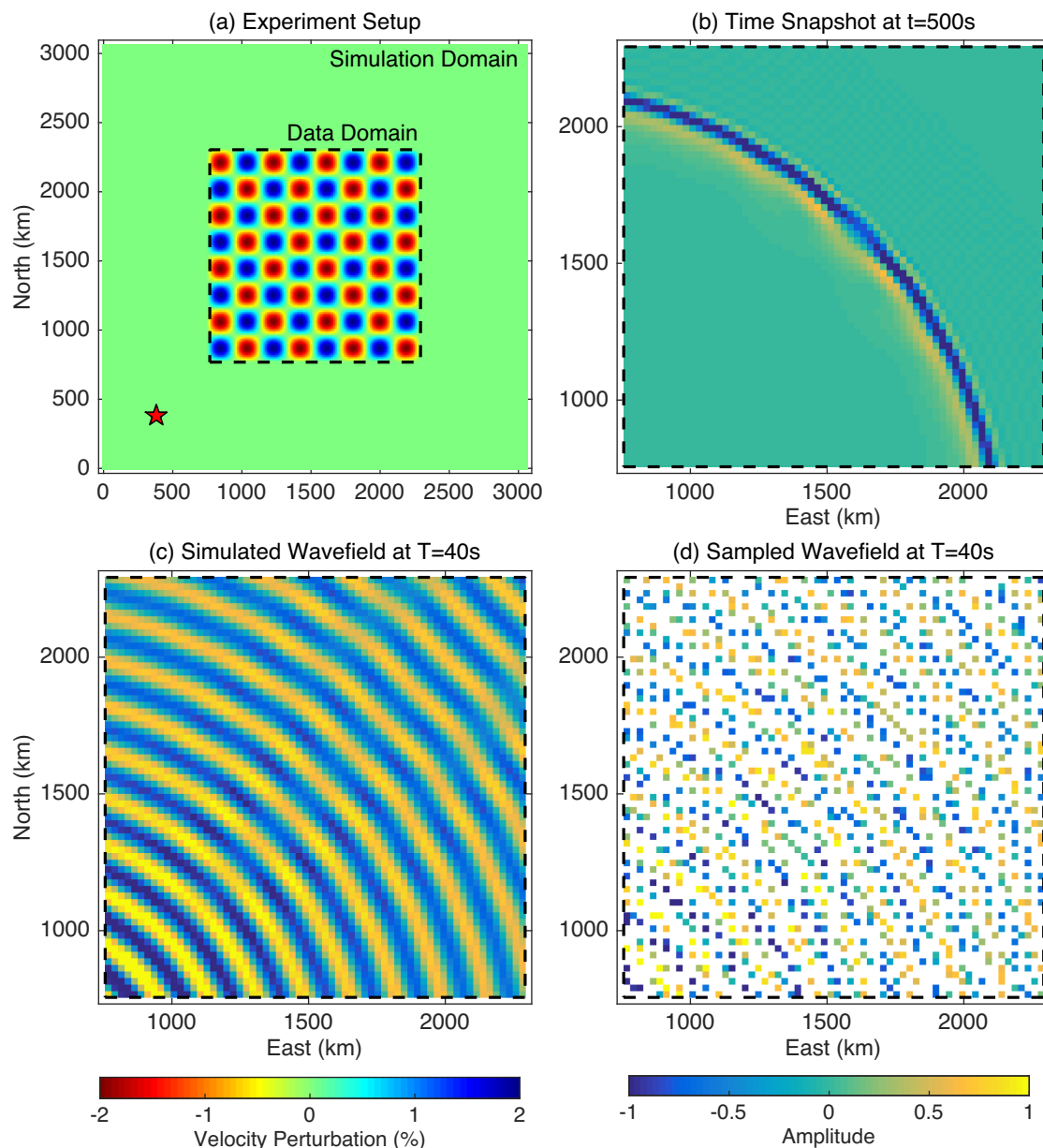


Figure 1. Model setup for synthetic tests. (a) The finite-difference simulation domain and the checkerboard velocity model used in the synthetic tests; scale and amplitude of velocity anomalies vary in different tests. The red star shows location of the source, and only synthetics within the dashed box, the data domain, are used in wavefield reconstruction. The grid spacing is uniformly 24 km. (b) A snapshot of the simulated wavefield within the data domain (dashed box) at time = 450 s. (c) The real part of complex-valued wavefield for the period of 40 s within the data domain (dashed box). (d) The sampled wavefield obtained by randomly removing 75 per cent of the gridpoints in (c) and adding independent white noise. Therefore, the grids are now irregular, and the average station spacing is 48 km.

$U(\vec{x})$ within the data domain for the period of 40 s is displayed in Fig. 1(c), on a 64×64 grid. The grid size here is 24 km, decimated from the 12 km FD simulation grid, but still approximately $1/7$ of the average wavelength for the 40-s-period Rayleigh waves. Therefore, the wavefield on this dense regular grid can be considered as the true wavefield. Fig. 1(d) is the ‘observed/sampled’ wavefield $U(\vec{x}_i)$ derived from the true wavefield by randomly removing 75 per cent of the gridpoints in Fig. 1(c) and adding independent white noise with an rms amplitude of 5 per cent of the average signal rms amplitude. Note that in this case the sample points are not truly random but have to be on the grid. However, the station spacing and distribution are random enough to demonstrate the effects

of random sampling. Our wavefield reconstruction method aims to best recover the missing 75 per cent part of the wavefield based on the observed 25 per cent part. The average station spacing is 48 km, still less than half of the average wavelength ($\lambda = 160$ km), thus satisfying the criteria of the Shannon–Nyquist sampling theorem.

Appropriate damping parameter λ in eq. (5) is important to avoid overdamping or overfitting. For the synthetic examples in this paper, it is relatively straightforward to choose λ by cross-validation. For real data, we may continue to use cross-validation or switch to the L-curve method instead. Fig. 2(a) shows both the classical L-curve (i.e. misfit of the observed 25 per cent versus model norm) and the

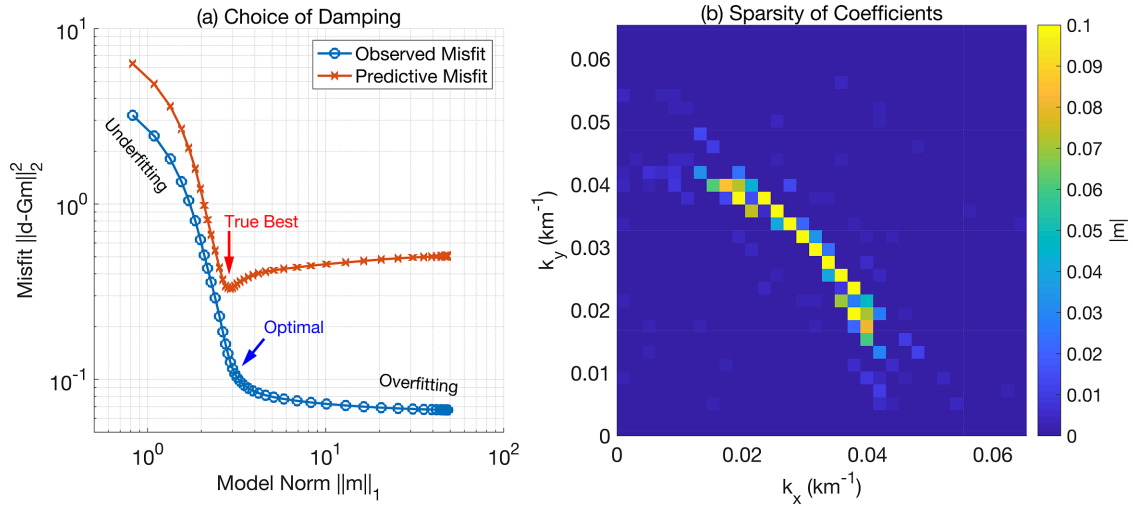


Figure 2. (a) Choice of damping parameter. The blue line is the L-curve of the L2 misfit of sampled wavefield (Fig. 1c) versus the L1 model norm in eq. (5). The clear corner in the L-curve, which is often considered as the optimal damping, coincides almost exactly with the minimum of the predictive misfit (for the missing 75 per cent data, red curve), implying consistent damping parameters favoured by the cross-validation and L-curve methods. (b) The distribution of plane-wave coefficients in the wavenumber domain. Most large coefficients concentrate around the average wavenumber of 0.039 km^{-1} for the 40-s period surface waves with an average of 4 km s^{-1} phase speed. Only 3 per cent of all the coefficients have absolute values larger than 0.01, which is 5 per cent of the largest coefficient.

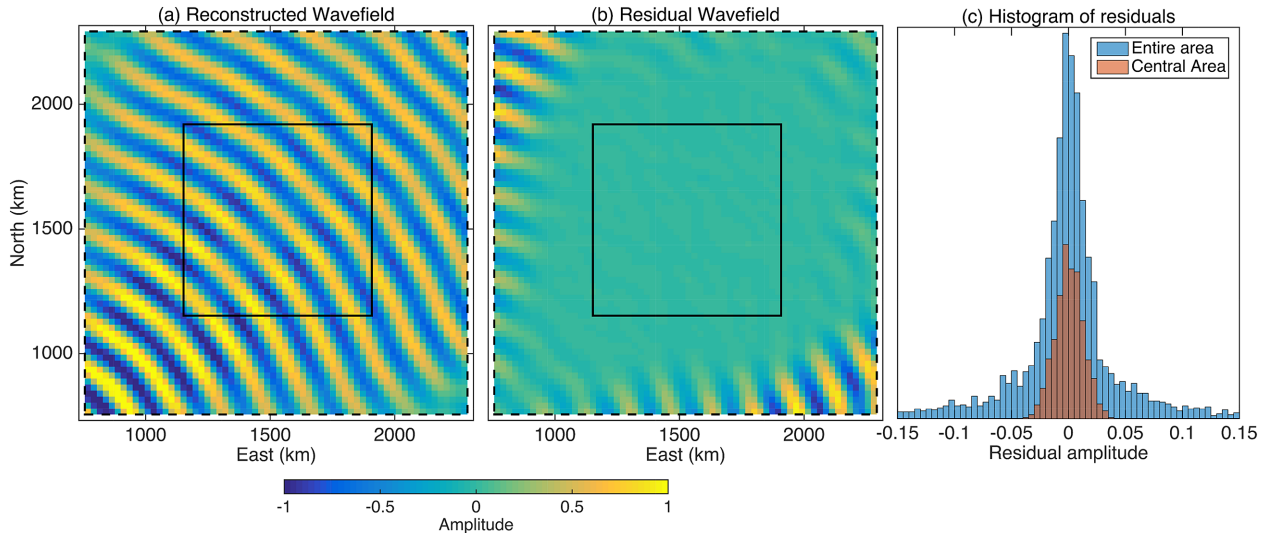


Figure 3. (a) The real part of the wavefield reconstructed from the sampled wavefield in Fig. 1(d), and (b) the residuals compared to the true wavefield (Fig. 1c). The residual is small in most areas except near the edges, due to the spatial weighting in eq. (5). Histograms in (c) show that residuals are less than 5 per cent of the average amplitude in the central square, while much larger outside due to the edge effect. In this paper, we only invert the reconstructed wavefield within the central square marked by the solid lines for velocity models.

cross-validation curve (i.e. predictive misfit of the missing 75 per cent versus model norm). The clear corner in the classical L-curve coincides almost exactly with the minimum of the predictive misfit, implying consistent damping parameters favoured by both L-curve and cross-validation methods. Using the optimal damping parameter, we invert for the plane-wave coefficients and reconstruct the wavefield on the same 64×64 grid as input (Fig. 3a). Fig. 2(b) shows the distribution of the resulted sparse plane-wave coefficients in the wavenumber domain. Fig. 3(b) displays the difference between the reconstructed wavefield and the true wavefield (Fig. 1c), and only near the edges of the data domain we observe substantial deviations (Fig. 3c). The poor reconstruction near edges is related to the spatial weighting term, w in eq. (5), as discussed above. In this paper, tomography techniques are only applied to the centre quarter

of the reconstructed wavefield (black square in Fig. 3) to avoid the edge effects. Using a spatially localized basis/frame (e.g. wavelet and curvelet) may help reduce or eliminate the edge effects in the future.

Once the wavefield $\tilde{U}(\vec{x}, \omega)$ is reconstructed for all interested frequencies, then \tilde{U} can be inverse Fourier transformed to the time domain, and we can apply the methods reviewed in Section 2 directly (Liang & Langston 2009; Lin & Ritzwoller 2011; de Ridder & Biondi 2015). However, due to the dispersive nature of surface waves, we prefer to operate directly on $\tilde{U}(\vec{x}, \omega)$ in the frequency domain, using the frequency-domain equivalents of the wavefield-based methods. Specifically in the following, we will apply the Helmholtz tomography (Lin & Ritzwoller 2011) and the WG (de Ridder & Biondi 2015).

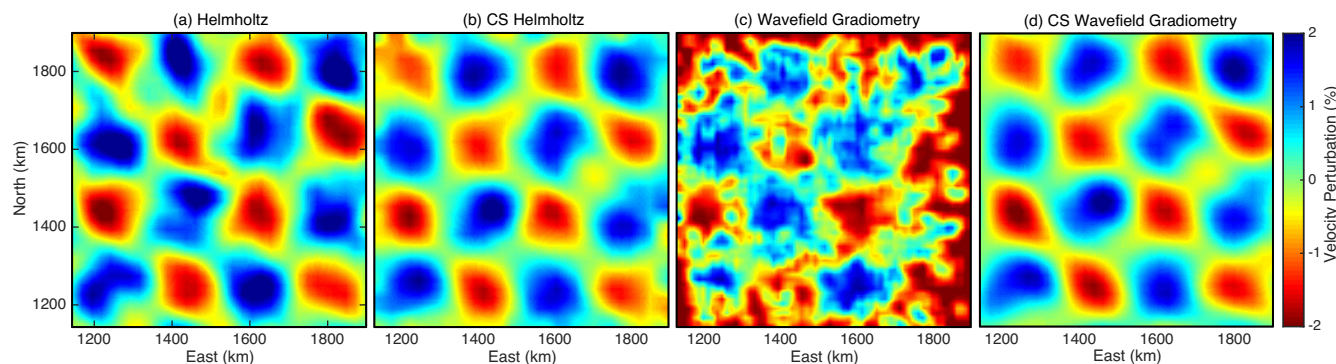


Figure 4. Velocity models estimated by applying Helmholtz tomography and wavefield gradiometry (WG) to the observed (Fig. 1d) and CS-reconstructed wavefield (Fig. 3a). The area displayed here is the centre square in Fig. 3.

To apply frequency-domain Helmholtz tomography, we need to derive amplitude and traveltime fields from $\tilde{U}(\vec{x}, \omega)$. While the amplitude field $A(\vec{r}) = |\tilde{U}(\vec{x}, \omega)|$ is straightforward to calculate, we need to unwrap the phase of $\tilde{U}(\vec{x}, \omega)$, between $-\pi$ and π , to a continuous field of traveltime $\tau(\vec{r}) + \tau_0$, where τ_0 is an arbitrary constant. Here, we use the classical Goldstein algorithm developed for Interferometric Synthetic Aperture Radar (InSAR) phase unwrapping (Goldstein *et al.* 1988). Phase reconstruction on grids with near-zero amplitudes may be unstable, so we apply a mask to the areas with amplitude smaller than 5 per cent of the average amplitude. This is similar to the common practices in InSAR data processing to mask areas of substantial decorrelation (e.g. vegetation and damage). In the last decade, more robust and efficient 2-D or 3-D phase unwrapping methods have been developed in the InSAR community (Chen & Zebker 2002; Bioucas-Dias & Valadao 2007; Hooper & Zebker 2007), which will likely improve our results, especially if the phase field is complicated. However, for most of the current synthetic tests, we find the Goldstein algorithm to be adequate. Figs 4(a) and (b) show the models recovered from traditional Helmholtz tomography and CS-Helmholtz tomography (i.e. frequency-domain Helmholtz tomography applied to the CS-reconstructed wavefield), respectively. Both models are very close to the true model, with normalized coefficients of 92 per cent and 95 per cent when cross-correlated with the true velocity model, respectively. Note that traditional Helmholtz tomography use a narrow-band filter in the time domain, while in CS-Helmholtz tomography, we average the results from individual frequency slices within the same band for the final image.

It proves inaccurate to apply WG directly to the irregularly spaced observations with noise (Fig. 4c), because irregular FD must be used to calculate the second-order spatial derivatives in eq. (3) and noise can be enhanced substantially. This difficulty of WG in evaluation of the spatial derivatives was also discussed by de Ridder & Curtis (2016) as a major challenge to the WG method. Fig. 4(c) shows the model estimated by WG using an irregular FD approximation based on Fornberg (1988), with additional smoothing afterward. Although the general pattern of fast and slow anomalies is consistent with the true model, the details are quite different, with a normalized cross-correlation coefficient of 55 per cent. However, if we apply WG to the CS-reconstructed wavefield on a dense regular grid, we recover a much better model, as shown in Fig. 4(d), with a normalized cross-correlation coefficient of 95 per cent. We attribute the great improvement to both the dense regular grid that enables better FD approximation, and the removal of the added incoherent noise that cannot be represented sparsely by plane waves. It is possible that another irregular FD approximation or calibration may substantially

improve (Fig. 4c). But essentially, we are proposing a new method to compute the FDs in WG by wavefield reconstruction followed by a regular-grid FD.

4 MORE SYNTHETIC EXAMPLES

In this section, we will demonstrate the CS of seismic wavefield using more synthetic examples. Specifically, we want to explore the performance of CS with sub-Nyquist sampling and complex wavefield.

4.1 Wavefield CS with sub-Nyquist sampling

One important advantage of wavefield CS is to reconstruct wavefield with fewer samples than required by classic sampling theory. In the example above, the average station spacing is still shorter than half of the average wavelength, therefore can be considered relatively dense. Here based on the same simulation setup as in Section 3 (except within an enhanced maximum velocity perturbations of ± 5 per cent), we reduce the target period from 40 to 30 s to generate a wavefield of shorter average wavelength, ~ 120 km (Fig. 5a). Then we retain only 12.5 per cent of the gridpoints randomly as the ‘observed’ wavefield, half of the samples in the test above, as shown in Fig. 5(b). The average station spacing is ~ 68 km, slightly larger than the 60 km spacing required by the Nyquist–Shannon sampling theorem. In fact, it becomes difficult to visually identify the propagation direction in Fig. 5(b), due to the aliasing effect. After choosing an appropriate damping parameter based on the L-curve approach (Fig. 5c), we reconstruct the wavefield as shown in Fig. 5(d), which is visually very similar to the true wavefield (Fig. 5a) and only deviates near the edges (Fig. 5e). Applying WG to the CS-reconstructed wavefield recovers the velocity model well in the centre quarter (Fig. 5f; normalized cross-correlation coefficient 97 per cent), suggesting a highly accurate reconstruction even though the sampled wavefield is aliased.

To further demonstrate wavefield CS in scenarios of sub-Nyquist sampling, we reduce the number of observational grids by another half, retaining only 6.25 per cent of the grids in Fig. 5(a). Thus, the sampled wavefield, as shown in Fig. 6(a), has an average station spacing of 96 km, substantially larger than the 60 km spacing required by the Nyquist–Shannon sampling theorem. The aliasing effect is so severe that it is impossible to visually identify wave propagation direction in Fig. 6(a). However, we are still able to reconstruct a full wavefield (Fig. 6b) that visually resembles the true wavefield (Fig. 5a), except near edges (Fig. 6c). This level of

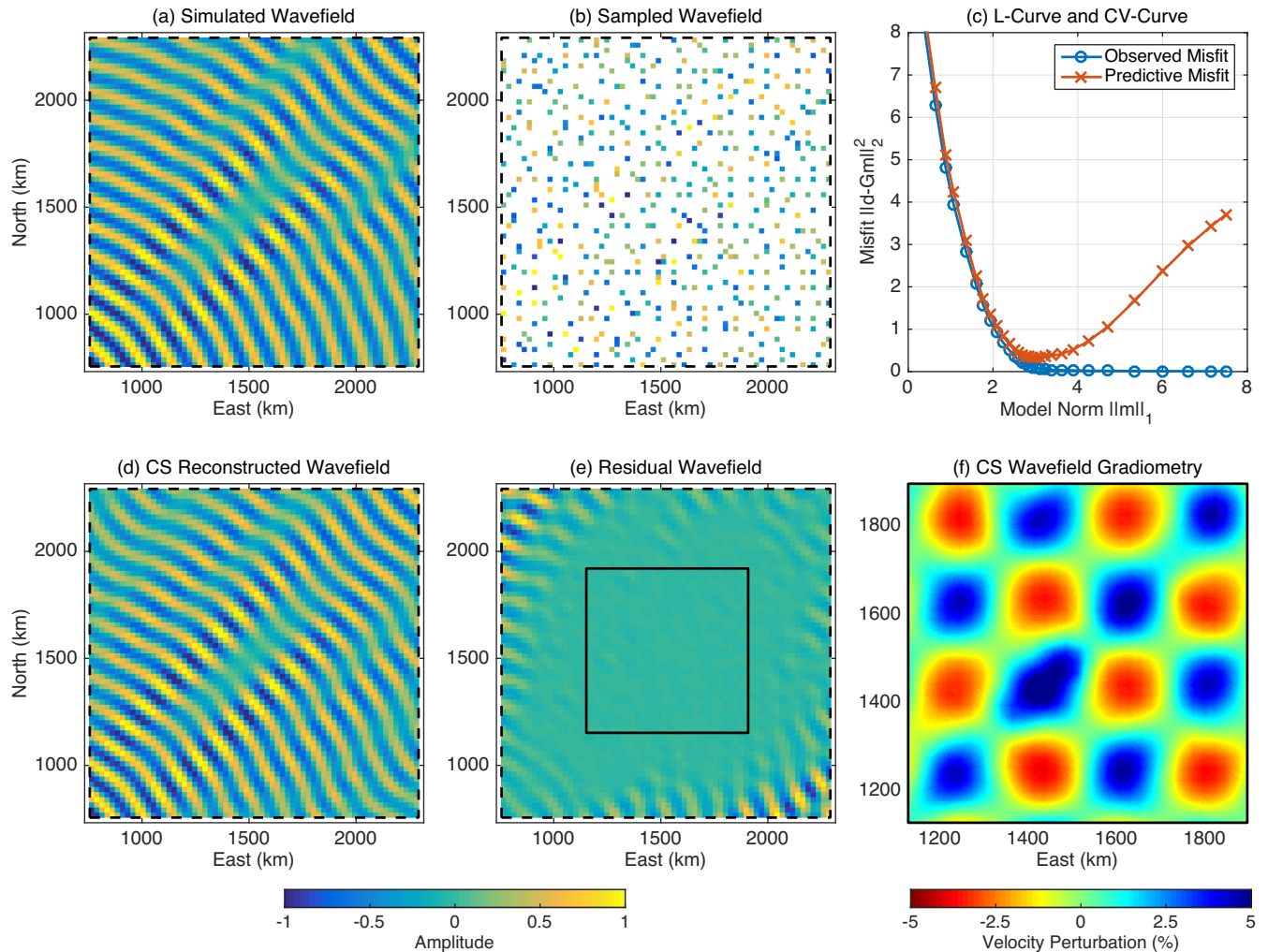


Figure 5. Wavefield CS for 30 s period with 12.5 per cent of the gridpoints. (a) and (b) show the simulated and sampled wavefields, respectively. The average station spacing is ~ 68 km, slightly larger than the 60 km spacing required by the Nyquist–Shannon theorem for the average wavelength of 120 km. (c) L-curve of observed misfit versus model norm, and cross-validation (CV) curve of predictive misfit versus model norm. The observed and predictive misfits are normalized by their sample sizes (12.5 per cent versus 87.5 per cent). We choose the damping parameter at the L-curve corner, which also matches well with the minimum of the predictive misfit. (d) and (e) display the reconstructed wavefield and its deviation from the true wavefield. (f) Velocity model estimated by applying wavefield gradiometry to the CS-reconstructed wavefield within the centre square of (e).

sub-Nyquist reconstructing accuracy is somewhat expected from the CS approach, but also suggests that our choice of plane-wave basis for surface wavefield of individual frequencies is at least appropriate (not necessarily optimal). The velocity model from applying WG to the CS-reconstructed wavefield reproduces well the overall patterns, but not the details or amplitudes (Fig. 6d; normalized cross-correlation coefficient 87 per cent). This is probably because some details of the wavefield are still missing after CS.

4.2 CS of complex wavefield

Here, we explore how complex wavefield may affect the reconstruction and tomography. One common cause of complex wavefield is multipathing over strong velocity heterogeneities. Here, we simulate the multipathing effect with the same setup as in Fig. 1(a), but with ± 15 per cent maximum velocity perturbations, which are stronger than perturbations in most long-period tomographic models but not uncommon in some recent short-period surface wave models (Ekström 2014, 2017). Fig. 7(a) is a snapshot of the simulated

wavefield, with clear multipathing as well as focusing/defocusing effects. In this case, traditional Helmholtz tomography in the time domain would face difficulties when measuring traveltimes by cross-correlating or phase tracking, because interferences of the multiple arrivals cause the waveforms to change rapidly. However, in CS–Helmholtz tomography, because ‘traveltime’ is derived by unwrapping phase in frequency domain, all arrivals’ contribution to the wavefield is taken into account appropriately.

After transferring the synthetics from the time to the frequency domains and retain 25 per cent of the gridpoints randomly (48 km average spacing), we get the sampled wavefield for the period of 30 s (Fig. 7b). We then reconstruct the full wavefield from the sampled wavefield, as shown in Fig. 7(c), which again only deviates from the true wavefield near the edges (Fig. 7d). Due to the multipathing effect, the wavefield has more complicated patterns than wavefields from models with weaker anomalies (Fig. 7c versus Fig. 1c). In particular, destructive interferences cause near-zero amplitude and unstable phase unwrapping in some areas. In CS–Helmholtz tomography, we mask these unstable areas before applying Goldstein phase unwrapping. Therefore, the CS–Helmholtz velocity model is

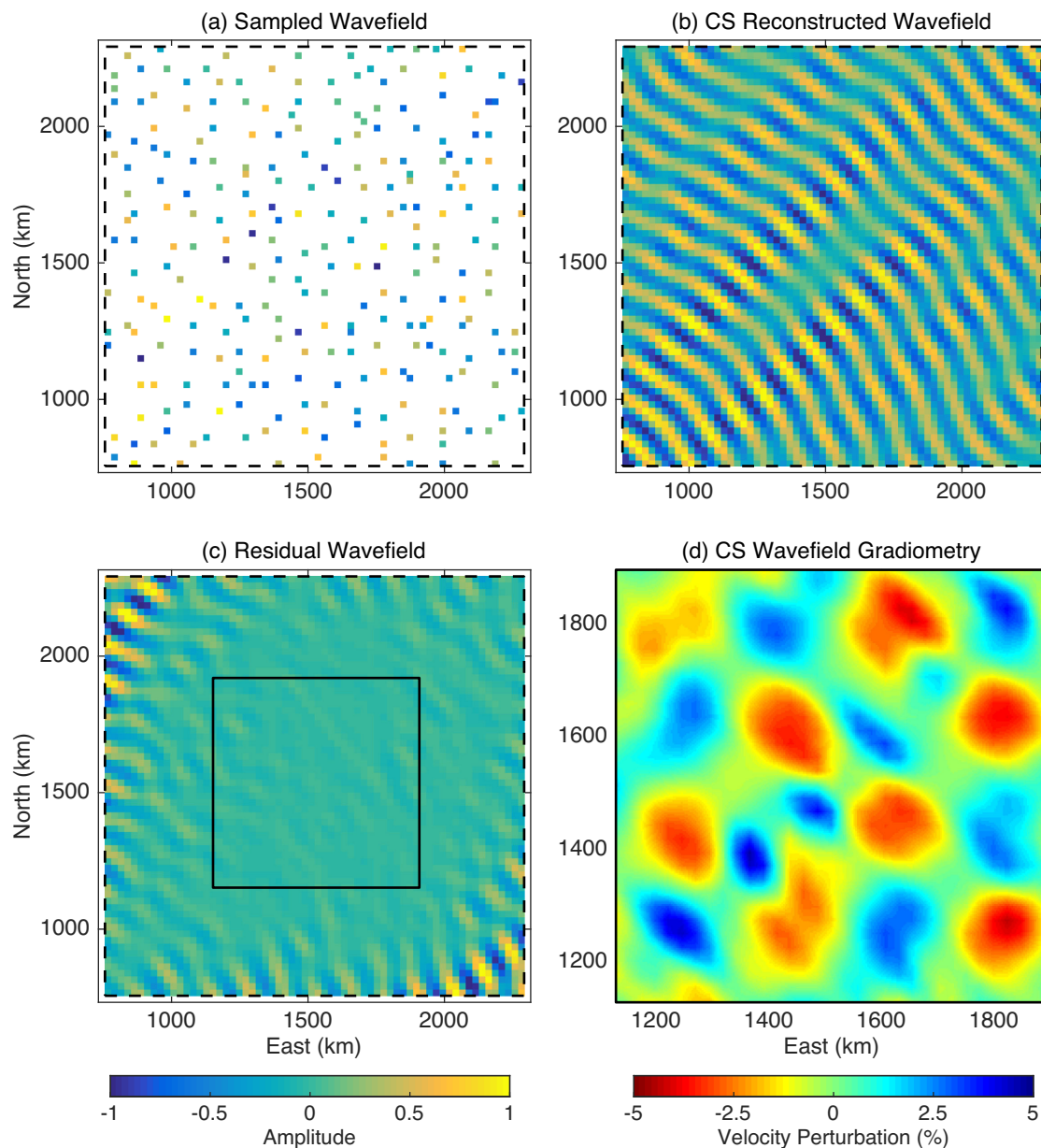


Figure 6. Similar to Fig. 5, but with only 6.25 per cent of the gridpoints used in reconstruction. The average station spacing is 96 km, significantly larger than required by the Nyquist–Shannon sampling theorem. (b) shows the wavefield reconstructed from the sampled wavefield in (a), and the difference from the true wavefield (Fig. 5a) is only substantial near edges (c). (d) displays the CS-WG velocity model estimated from the reconstructed wavefield.

incomplete, even though in most areas the velocity perturbations are well recovered (Fig. 7e; normalized cross-correlation coefficient 86 per cent). CS–WG has no difficulty in dealing with wave interferences and recovers the complete velocity model well, as shown in Fig. 7(f) with a normalized cross-correlation coefficient of 94 per cent.

Another cause of wavefield complexity is interference of waves from multiple simultaneous sources, such as a large earthquake with extended rupture and duration. An extreme example of this kind of complexity is the microseism noise wavefield, which is excited by distributed continuous sources in the ocean or near the coast. For this kind of complexity, traditional Helmholtz tomography has trouble in measuring amplitudes and traveltimes. de Ridder & Biondi (2015) applied WG directly to the ambient noise field on a dense array and retrieved high-resolution velocity models without

cross-correlating long durations of noise records. In Section 3, we demonstrated that CS helps improve WG resolution and robustness by first reconstructing the full wavefield. Here, we test if CS still recovers wavefield accurately with interfering waves.

Fig. 8(a) shows a synthetic wavefield snapshot, due to five sources of random amplitudes distributed to the southwest corner of the simulation domain. Here, the velocity perturbation is ± 5 per cent, the same as in Section 4.1 and Fig. 5, so the complex wavefield patterns are mostly due to the five interfering wave fronts. The real part of the wavefield for a single period (30 s) is displayed in Fig. 8(b), after removing 75 per cent of the grids. The average station spacing is 48 km. Fig. 8(c) shows the CS-reconstructed wavefield, which is only substantially different from the true wavefield near edges (Fig. 8d). This suggests that, at least for a small number of simultaneous sources, CS is still able to recover the full wavefield accurately

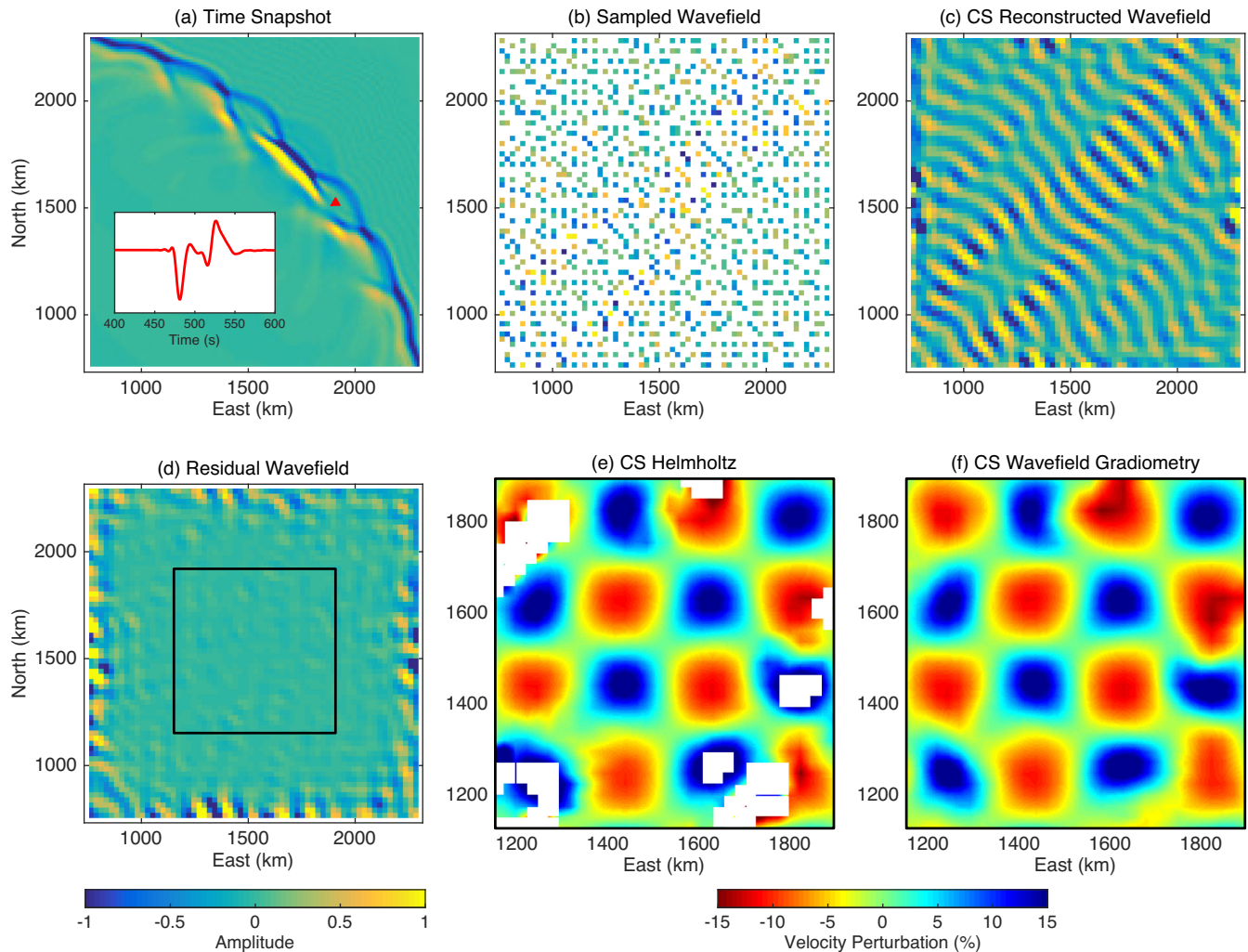


Figure 7. Synthetic test with complex wavefield due to strong multipathing. (a) This shows a snapshot of the broad-band wavefield traveling through a checkerboard velocity model with ± 15 per cent maximum velocity perturbations. Inset shows the synthetic seismogram at the red triangle, with two multipathing arrivals. The real part of the wavefield for the period of 30 s is displayed in (b), after 75 per cent of the grids being removed. The average station spacing is 48 km. (c) The CS-reconstructed wavefield, displaying complex patterns due to multipathing. (d) Difference of the reconstructed wavefield from the true wavefield. The centre square marked by solid lines is where velocity model has been estimated using Helmholtz tomography (e) and wavefield gradiometry (f). The blank areas in (e) are caused by masks to avoid unstable phase unwrapping for gridpoints with near-zero amplitudes.

under the sparsity assumption. Applying Helmholtz tomography and WG to the reconstructed wavefield both produce good velocity models (Figs 8e and f; normalized cross-correlation coefficients 95 per cent and 97 per cent, respectively), which is somewhat surprising considering that similar level of wavefield complexity in Fig. 7 caused difficulties in phase unwrapping in certain areas. This may be related to the fact that wave interference is more strongly frequency dependent than multipathing within a narrow band.

4.3 CS to recover small-scale heterogeneities

In wavefield-based imaging/tomography, spatial resolution is controlled by both seismic wavelength and station density. In this last example, we aim to test how well CS can preserve information within wavefield on small-scale heterogeneities, and whether better resolution is possible over traditional methods. The example setup is the same as in previous tests, but each high- or low-velocity block in the input checkerboard model now has a spatial dimension of $128 \text{ km} \times 128 \text{ km}$, smaller than the $192 \text{ km} \times 192 \text{ km}$ in previous

models. Our target period is 45 s and the average seismic wavelength is 180 km. We then retain only 9 per cent of the gridpoints to produce an average station spacing of 80 km (Fig. 9a), smaller than half of the average wavelength but larger than half of the block size. This means that our sampling is super-Nyquist for the waves but the spatial pattern of velocity model is not well sampled (Fig. 9c). Therefore, if we apply traditional Helmholtz tomography to the wavefield samples, the commonly used minimum-curvature smoothing of traveltimes and amplitudes cannot preserve details of the fields and will produce distorted velocity model with smeared blocks and underestimated anomalies (Fig. 9d; normalized cross-correlation coefficient 78 per cent). On the other hand, if we apply Helmholtz tomography and WG to the CS-reconstructed wavefield (Fig. 9b), we recover well the spatial patterns of velocity model, even though the amplitude of anomaly is still underestimated (Figs 9e and f; normalized cross-correlation coefficients 90 per cent). This suggests that the assumptions behind wavefield CS actually help better preserve details of the wavefield about small-scale heterogeneities, hence improve resolution of tomography.

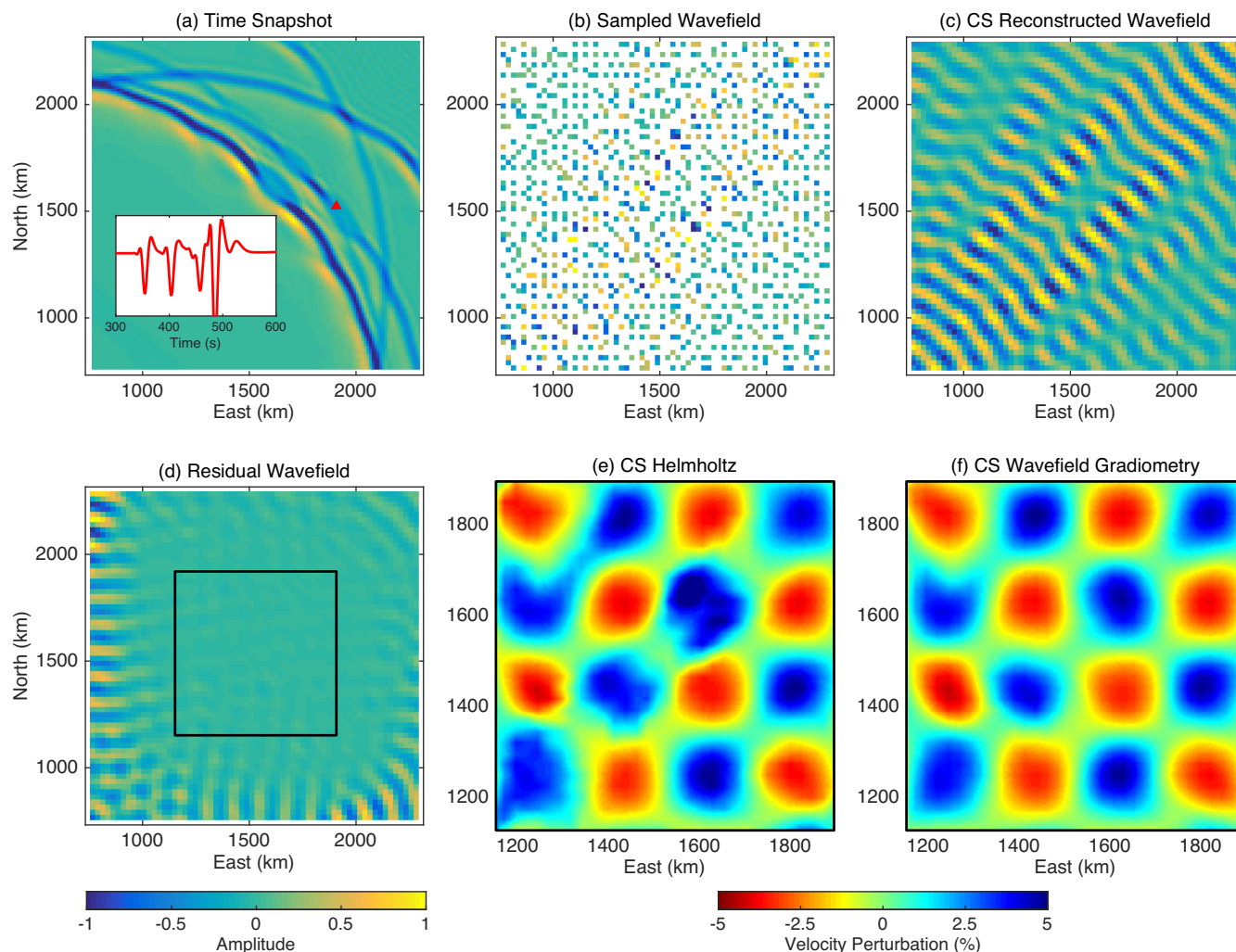


Figure 8. Similar to Fig. 7, but for a synthetic test with ± 5 per cent velocity perturbations, yet a complex wavefield caused by interferences of waves from five simultaneous sources in the southwest corner of the simulation domain.

5 DISCUSSION

The synthetic examples above show the advantages and disadvantages of the Helmholtz tomography and WG. Operating on the traveltimes and amplitude fields in Helmholtz tomography makes it more robust to random noise, and the minimum-curvature smoothing works remarkably well when the target velocity model has only weak gradients. In contrast, WG makes less assumptions so it applies well to much more complicated wavefield, either due to very heterogeneous structures or complex sources. However, WG is very sensitive to noise with its second-order spatial derivatives directly on irregular samples.

Our CS method helps improve both methods. For Helmholtz tomography in complex wavefield, CS works in the frequency domain and helps take into account wave interferences appropriately. We measure the ‘traveltimes’ field by unwrapping the phase of reconstructed wavefield, which proves much more robust and accurate than measurements in the time domain for a complex wavefield. Future replacement of the Goldstein algorithm with more advanced phase unwrapping algorithms, especially the ones that combine multiple frequencies, should further improve the accuracy and robustness. For WG, CS helps remove significant numerical errors due to irregular FDs and random noise. In some sense, CS WG makes relatively few assumptions and applies to the most general cases of

surface wave tomography, even for a very complex wavefield. More advanced WG that considers anisotropy and attenuation should also be able to benefit from the CS wavefield reconstruction.

One key requirement of CS is incoherence between the sampling function and the representing basis/frame. In this study, incoherence is guaranteed by the random spatial sampling of the wavefields. If we choose a regular sampling with sub-Nyquist station spacing, it is impossible to recover wavefield accurately according to the Nyquist–Shannon sampling theorem. On the other hand, not all different kinds of random sampling are equally effective. In the tests discussed above, the random sampling is relatively uniform without obvious clustering of stations or large gaps in data. For a given array geometry [e.g. the Southern California Seismic Network (SCSN) or the USArray], synthetic tests will be necessary to investigate its resolving power at different frequencies. For future seismic arrays, optimal design of the sampling regime for a particular problem may help reduce acquisition cost and improve accuracy and is worthy of further investigation (Allegor *et al.* 2017; Baraniuk & Steeghs 2017).

Another benefit of the CS approach is that it can efficiently reduce incoherent noise, which is not sparse under certain wave bases/frames (e.g. plane waves and curvelets). This feature is not demonstrated in depth in this paper, but CS has been a

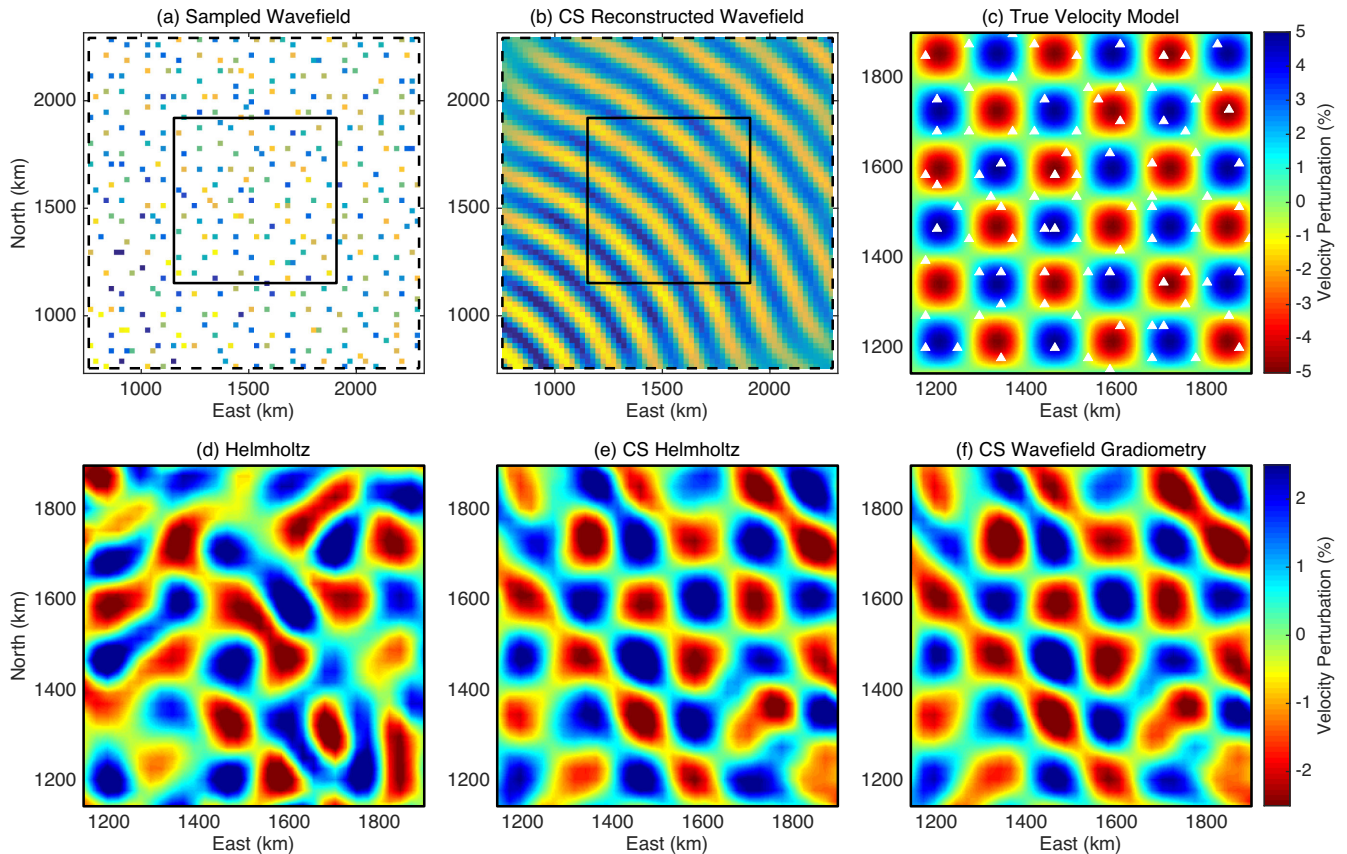


Figure 9. Synthetic test on wavefield CS's sensitivity to small-scale heterogeneities. (a) Sampled wavefield at 45 s period with an average station spacing of 80 km, and (b) the reconstructed wavefield. (c) Input velocity model within the centre square in (a) and (b), with stations shown in white triangles. The block size of velocity anomalies is 128 km \times 128 km. (d)–(f) display the recovered velocity models from traditional Helmholtz tomography applied to the sampled wavefield, Helmholtz tomography and WG applied to the CS-reconstructed wavefield, respectively.

popular denoising approach in exploration seismology (Hennenfent & Herrmann 2006; Herrmann *et al.* 2007). Noise and gaps in data will introduce uncertainties to the reconstructed wavefield, which are difficult to quantify with our current convex optimization methodology (i.e. TFOCS). A new class of method called Bayesian CS (Ji *et al.* 2008), although computationally more expensive, can provide posterior probability distribution function that can be used to identify areas with large uncertainties. Therefore in the future, we may be able to mask areas with large uncertainties to avoid overinterpreting parts of the reconstructed wavefield that are corrupted by noise or not well sampled.

6 CONCLUSIONS

In this paper, we have presented a new wavefield reconstruction method based on the theory of compressing sensing. We make a fundamental assumption that 2-D surface wavefield is optimally sparse under the plane-wave representation, that is, sparse in the wavenumber domain. We choose a combination of the plane-wave basis and sparsity-promoting ℓ_1 optimization in reconstructing the complete wavefield from irregularly sampled and noisy data. We apply the method to single-frequency surface waves sampled by 2-D dense arrays. Helmholtz tomography or WG applied to the reconstructed wavefield delivers improved velocity models compared with the ones from the same methods applied to the original wavefield. The improvements are especially obvious when sampling is sparser than

required by the Nyquist–Shannon sampling theorem, or when wavefield is complicated by multipathing/scattering/distributed sources.

The CS idea is not limited to 2-D surface wave tomography. Instead, it is a more general scheme about how to best reconstruct the seismic wavefield with dense arrays. We can potentially apply backprojection, earthquake detection, reverse time migration, or any other wavefield-based method to the reconstructed wavefield. In many of these other cases, we will be working on body waves, which are solutions to 3-D wave equation but recorded on 2-D free surface. Therefore, we may need to operate in the time domain and extend to 3-D basis/frame, or use ensemble Kalman filter to connect 2-D snapshots of the wavefield (Hoshiba & Aoki 2015). Similar challenges exist in InSAR data processing, which has spatially dense and uniform samples for a large area but irregular and sparse samples in time. How to reconstruct the full spatiotemporal deformation field (at different time scales) is a common challenge for seismology and geodesy (Riel *et al.* 2014).

ACKNOWLEDGEMENTS

We thank Peter Gerstoft, Victor Tsai, Jack Muir and Mark Simons for helpful discussions. Carl Tape and another anonymous reviewer's comments help clarify the paper. ZZ is partly supported by the NSF 1722879. QL thanks the China Scholarship Council for sponsoring his study at California Institute of Technology. QL and JH are supported by NSFC (41274124), Chinese National 973 Project 2014CB239006 and Major National Oil and Gas Projects

(2016ZX05014-001-008 and 2016ZX05002005-007). There are no real data involved in this study.

REFERENCES

- Allegar, N., Herrmann, F.J. & Mosher, C.C., 2017. Introduction to this special section: impact of compressive sensing on seismic data acquisition and processing, *Leading Edge*, **36**(8), 640–640.
- Bao, X., Dalton, C.A., Jin, G., Gaherty, J.B. & Shen, Y., 2016. Imaging Rayleigh wave attenuation with USArray, *Geophys. J. Int.*, **206**(1), 241–259.
- Baraniuk, R.G. & Steeghs, P., 2017. Compressive sensing: a new approach to seismic data acquisition, *Leading Edge*, **36**(8), 642–645.
- Becker, S.R., Candès, E.J. & Grant, M.C., 2011. Templates for convex cone problems with applications to sparse signal recovery, *Math. Prog. Comp.*, **3**(3), 165–218.
- Ben-Zion, Y. *et al.*, 2015. Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone, *Geophys. J. Int.*, **202**(1), 370–380.
- Bioucas-Dias, J.M. & Valadao, G., 2007. Phase unwrapping via graph cuts, *IEEE Trans. Image Process.*, **16**(3), 698–709.
- Bobin, J., Starck, J.-L. & Ottensmeyer, R., 2008. Compressed sensing in astronomy, *IEEE J. Sel. Top. Signal Process.*, **2**(5), 718–726.
- Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.*, **117**, B02301, doi:10.1029/2011JB008560.
- Boehm, C., Hanzlich, M., de la Puente, J. & Fichtner, A., 2016. Wavefield compression for adjoint methods in full-waveform inversion, *Geophysics*, **81**(6), R385–R397.
- Candès, E.J. & Demanet, L., 2005. The curvelet representation of wave propagators is optimally sparse, *Commun. Pure appl. Math.*, **58**(11), 1472–1528.
- Candès, E.J., Demanet, L., Donoho, D. & Ying, L., 2006. Fast discrete curvelet transforms, *Multiscale Model. Simul.*, **5**(3), 861–899.
- Candès, E.J. & Wakin, M.B., 2008. An introduction to compressive sampling, *IEEE Signal Process. Mag.*, **25**(2), 21–30.
- Candès, E.J., Wakin, M.B. & Boyd, S.P., 2008. Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.*, **14**(5–6), 877–905.
- Chen, C.W. & Zebker, H.A., 2002. Phase unwrapping for large SAR interferograms: statistical segmentation and generalized network models, *IEEE Trans. Geosci. Remote Sens.*, **40**(8), 1709–1719.
- Davenport, M.A., Duarte, M.F., Eldar, Y.C. & Kutyniok, G., 2011. Introduction to compressed sensing, in *Compressed Sensing: Theory and Applications*, pp. 1–68, Cambridge, U.K.: Cambridge Univ. Press.
- de Ridder, S. & Biondi, B., 2015. Near-surface Scholte wave velocities at Ekofisk from short noise recordings by seismic noise gradiometry, *Geophys. Res. Lett.*, **42**(17), 7031–7038.
- de Ridder, S. & Curtis, A., 2016. Anisotropic seismic-noise gradiometry with finite-difference stencil correction, in *SEG Technical Program Expanded Abstracts 2016*, edited, pp. 2699–2703. Society of Exploration Geophysicists.
- Demanet, L. & Ying, L., 2007. Wave atoms and sparsity of oscillatory patterns, *Appl. Comput. Harmon. Anal.*, **23**(3), 368–387.
- Donoho, D.L., 2006. For most large underdetermined systems of linear equations the minimal L_1 -Norm solution is also the sparsest solution, *Commun. Pure appl. Math.*, **59**(6), 797–829.
- Duarte, M.F., Davenport, M.A., Takhar, D., Laska, J.N., Sun, T., Kelly, K.E. & Baraniuk, R.G., 2008. Single-pixel imaging via compressive sampling, *IEEE Signal Process. Mag.*, **25**(2), 83–91.
- Ekström, G., 2014. Love and rayleigh Phase-velocity maps, 5–40 s, of the western and central USA from USArray data, *Earth planet. Sci. Lett.*, **402**, 42–49.
- Ekström, G., 2017. Short-period surface-wave phase velocities across the conterminous United States, *Phys. Earth planet. Inter.*, **270**, 168–175.
- Fomel, S. & Liu, Y., 2010. Seislet transform and seislet frame, *Geophysics*, **75**(3), V25–V38.
- Fornberg, B., 1988. Generation of finite difference formulas on arbitrarily spaced grids, *Math. Comp.*, **51**(184), 699–699.
- Forsyth, D.W., Webb, S.C., Dorman, L.M. & Shen, Y., 1998. Phase velocities of Rayleigh waves in the MELT experiment on the East Pacific Rise, *Science*, **280**(5367), 1235–1238.
- Goldstein, R.M., Zebker, H.A. & Werner, C.L., 1988. Satellite radar interferometry: two-dimensional phase unwrapping, *Radio Sci.*, **23**(4), 713–720.
- Haney, M.M. & Tsai, V.C., 2015. Nonperturbational surface-wave inversion: a Dix-type relation for surface waves, *Geophysics*, **80**(6), EN167–EN177.
- Hansen, P.C. & O’Leary, D.P., 1993. The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J. Sci. Comput.*, **14**(6), 1487–1503.
- Hansen, S.M. & Schmandt, B., 2015. Automated detection and location of microseismicity at Mount St. Helens with a large-N geophone array, *Geophys. Res. Lett.*, **42**(18), 7390–7397.
- Hennenfent, G., Fenelon, L. & Herrmann, F.J., 2010. Nonequispaced curvelet transform for seismic data reconstruction: a sparsity-promoting approach, *Geophysics*, **75**(6), WB203–WB210.
- Hennenfent, G. & Herrmann, F.J., 2006. Seismic denoising with nonuniformly sampled curvelets, *Comput. Sci. Eng.*, **8**(3), 16–25.
- Herman, M.A. & Strohmer, T., 2009. High-resolution radar via compressed sensing, *IEEE Trans. Signal Process.*, **57**(6), 2275–2284.
- Herrmann, F.J. & Hennenfent, G., 2008. Non-parametric seismic data recovery with curvelet frames, *Geophys. J. Int.*, **173**(1), 233–248.
- Herrmann, F.J., Wang, D., Hennenfent, G. & Moghaddam, P.P., 2008. Curvelet-based seismic data processing: a multiscale and nonlinear approach, *Geophysics*, **73**(1), A1–A5.
- Hindriks, K. & Duijndam, A.J., 1999. Reconstruction of 3-D seismic signals irregularly sampled along two spatial coordinates, *Geophysics*, **65**(1), 253–263.
- Hooper, A. & Zebker, H.A., 2007. Phase unwrapping in three dimensions with application to InSAR time series, *J. Opt. Soc. Am. A*, **24**(9), 253–263.
- Hoshiba, M. & Aoki, S., 2015. Numerical Shake prediction for Earthquake early warning: data assimilation, real-time shake mapping, and simulation of wave propagation, *Bull. seism. Soc. Am.*, **105**(3), 1324–1338.
- Inbal, A., Ampuero, J.P. & Clayton, R.W., 2016. Localized seismic deformation in the upper mantle revealed by dense seismic arrays, *Science*, **354**(6308), 88–92.
- Ishii, M., Shearer, P.M., Houston, H. & Vidale, J.E., 2005. Extent, duration and speed of the 2004 Sumatra–Andaman earthquake imaged by the Hi-Net array, *Nature*, **435**(7044), 933–936.
- Ji, S., Xue, Y. & Carin, L., 2008. Bayesian compressive sensing, *IEEE Trans. Signal Process.*, **56**(6), 2346–2356.
- Jin, G. & Gaherty, J.B., 2015. Surface wave phase-velocity tomography based on multichannel cross-correlation, *Geophys. J. Int.*, **201**(3), 1383–1398.
- Julia, J., Ammon, C., Herrmann, R. & Correig, A.M., 2000. Joint inversion of receiver function and surface wave dispersion observations, *Geophys. J. Int.*, **143**(1), 99–112.
- Kumar, R., Da Silva, C., Akalin, O., Aravkin, A.Y., Mansour, H., Recht, B. & Herrmann, F.J., 2015. Efficient matrix completion for seismic data reconstruction, *Geophysics*, **80**(5), V97–V114.
- Kumar, R., Wason, H., Sharan, S. & Herrmann, F.J., 2017. Highly repeatable 3D compressive full-azimuth towed-streamer time-lapse acquisition — a numerical feasibility study at scale, *Leading Edge*, **36**(8), 677–687.
- Langston, C.A., 2007. Wave Gradiometry in the time domain, *Bull. seism. Soc. Am.*, **97**(3), 926–933.
- Langston, C.A., 2007. Wave Gradiometry in two dimensions, *Bull. seism. Soc. Am.*, **97**(2), 401–416.
- Liang, C. & Langston, C.A., 2009. Wave gradiometry for USArray: Rayleigh waves, *J. geophys. Res.*, **114**(B2), doi:10.1029/2008jb005918.
- Lin, F.-C., Li, D., Clayton, R.W. & Hollis, D., 2013. High-resolution 3D shallow crustal structure in long beach, California: application of ambient noise tomography on a dense seismic array, *Geophysics*, **78**(4), Q45–Q56.
- Lin, F.-C. & Ritzwoller, M.H., 2011. Helmholtz surface wave tomography for isotropic and azimuthally anisotropic structure, *Geophys. J. Int.*, **186**(3), 1104–1120.

- Lin, F.-C., Ritzwoller, M.H. & Snieder, R., 2009. Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array, *Geophys. J. Int.*, **177**(3), 1091–1110.
- Lin, F.-C., Tsai, V.C. & Ritzwoller, M.H., 2012. The local amplification of surface waves: a new observable to constrain elastic velocities, density, and anelastic attenuation, *J. geophys. Res.*, **117**, B06302, doi:10.1029/2012JB009208.
- Liu, Y. & Holt, W.E., 2015. Wave gradiometry and its link with Helmholtz equation solutions applied to USArray in the eastern US, *J. geophys. Res.*, **120**(8), 5717–5746.
- Lustig, M., Donoho, D. & Pauly, J.M., 2007. Sparse MRI: the application of compressed sensing for rapid MR imaging, *Magn. Reson. Med.*, **58**(6), 1182–1195.
- Riel, B., Simons, M., Agram, P. & Zhan, Z., 2014. Detecting transient signals in geodetic time series using sparse estimation techniques, *J. geophys. Res.*, **119**(6), 5140–5160.
- Ritzwoller, M., Shapiro, N. & Leahy, G., 2003. A resolved mantle anomaly as the cause of the Australian–Antarctic Discordance, *J. geophys. Res.*, **108**(B12), 2559, doi:10.1029/2003JB002522.
- Ritzwoller, M.H., Shapiro, N.M., Levshin, A.L. & Leahy, G.M., 2001. Crustal and upper mantle structure beneath Antarctica and surrounding oceans, *J. geophys. Res.*, **106**(B12), 30 645–30 670.
- Shen, W., Ritzwoller, M.H., Schulte-Pelkum, V. & Lin, F.-C., 2012. Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach, *Geophys. J. Int.*, **192**(2), 807–836.
- Tanimoto, T., 1990. Modelling curved surface wave paths: membrane surface wave synthetics, *Geophys. J. Int.*, **102**(1), 89–100.
- Tromp, J. & Dahlen, F., 1992. Variational principles for surface wave propagation on a laterally heterogeneous Earth–II. Frequency-domain JWKB theory, *Geophys. J. Int.*, **109**(3), 599–619.
- Tromp, J. & Dahlen, F., 1993. Variational principles for surface wave propagation on a laterally heterogeneous Earth–III. Potential representation, *Geophys. J. Int.*, **112**(2), 195–209.
- Wielandt, E., 1993. Propagation and structural interpretation of non-plane waves, *Geophys. J. Int.*, **113**(1), 45–53.
- Yang, Y. & Forsyth, D.W., 2006. Rayleigh wave phase velocities, small-scale convection, and azimuthal anisotropy beneath southern California, *J. geophys. Res.*, **111**(B7), doi:10.1029/2005jb004180.
- Yang, Y., Ritzwoller, M.H., Lin, F.-C., Moschetti, M. & Shapiro, N.M., 2008. Structure of the crust and uppermost mantle beneath the western United States revealed by ambient noise and earthquake tomography, *J. geophys. Res.*, **113**, B12310, doi:10.1029/2008JB005833.
- Yao, H., Gerstoft, P., Shearer, P.M. & Mecklenbräuker, C., 2011. Compressive sensing of the Tohoku–Oki Mw 9.0 earthquake: frequency-dependent rupture modes, *Geophys. Res. Lett.*, **38**, L20310, doi:10.1029/2011GL049223.
- Zwartjes, P. & Sacchi, M., 2007. Fourier reconstruction of nonuniformly sampled, aliased seismic data, *Geophysics*, **72**(1), V21–V32.